Review: to be or not to be an identifiable model. Is this a relevant question in animal science modelling?

Rafael Munoz Tamayo, Laurence Puillet, J. Daniel, Daniel Sauvant, O. Martin, Masoomeh Taghipoor, P. Blavy

To be or not to be an identifiable model. Is this a relevant question in animal science modelling?

Rafael Muñoz-Tamayo [1], Laurence Puillet [1], Jean-Baptiste Daniel [1,2], Daniel Sauvant [1], Olivier Martin[1], Masoomeh Taghipoor [3], and Pierre Blavy [1]

[1] *UMR Modélisation Systémique Appliquée aux Ruminants, INRA, AgroParisTech, Université Paris-Saclay, 75005, Paris, France*

[2]*Trouw Nutrition R&D, P.O. Box 220, 5830 AE Boxmeer, the Netherlands*

[3] *PEGASE, AgroCampus Ouest, INRA, 35590, Saint-Gilles, France*

Corresponding author: Rafael Muñoz-Tamayo.

Email: rafael.munoztamayo@agroparistech.fr

Short title: Animal modelling meets structural identifiability

Abstract

What is a good (useful) mathematical model in animal science? For models constructed for prediction purposes, the question of model adequacy (usefulness) has been traditionally tackled by statistical analysis applied to observed experimental data relative to model-predicted variables. However, little attention has been paid to analytic tools that exploit the mathematical properties of the model equations. For example, in the context of model calibration, before attempting a numerical estimation of the model parameters, we might want to know if we have any chance of success in estimating a unique best value of the model parameters from available measurements. This question of uniqueness is referred to as structural identifiability; a mathematical property that is defined on the sole basis of the model structure

within a hypothetical ideal experiment determined by a setting of model inputs (stimuli) and observable variables (measurements). Structural identifiability analysis applied to dynamic models described by ordinary differential equations (ODE) is a common practice in control engineering and system identification. This analysis demands mathematical technicalities that are beyond the academic background of animal science, which might explain the lack of pervasiveness of identifiability analysis in animal science modelling. To fill this gap, in this paper we address the analysis of structural identifiability from a practitioner perspective by capitalizing on the use of dedicated software tools. Our objectives are (i) to provide a comprehensive explanation of the structural identifiability notion for the community of animal science modelling, (ii) to assess the relevance of identifiability analysis in animal science modelling and (iii) to motivate the community to use identifiability analysis in the modelling practice (when the identifiability question is relevant). We focus our study on ODE models. By using illustrative examples that include published mathematical models describing lactation in cattle, we show how structural identifiability analysis can contribute to advancing mathematical modelling in animal science towards the production of useful models and highly informative experiments. Rather than attempting to impose a systematic identifiability analysis to the modelling community during model developments, we wish to open a window towards the discovery of a powerful tool for model construction and experiment design.

Keywords: dynamic modelling; identifiability; model calibration; optimal experiment design; parameter identification

Implications

Mathematical modelling has played a central role in animal science with a plethora of developments for enhancing understanding and guiding sustainable livestock farming. Progress in precision farming and omics technologies will call for model developments adapted to get the most out of the resulting big data, including better modelling practice. Our objective is of providing insight into a mathematical tool called structural identifiability analysis that has been seldom used for analysing dynamic models in animal science. We illustrate how this tool (when relevant) can contribute to advancing mathematical modelling towards the production of useful models.

Introduction

The development of mathematical models in animal science has contributed to gaining insight in different central aspects of animal physiology such as metabolism and digestion. The potential of modelling has been discussed by different authors (France, 1988; Baldwin, 2000; Doeschl-Wilson, 2011).

A classical modelling approach for describing the dynamics of a system under study is to construct dynamic models consisting of ODEs. These models comprise parameters (sometimes in large number) whose numerical values need to be estimated from experimental data by an adequate calibration routine. In animal science modelling, it is a common practice to assess model adequacy by statistical analysis applied on observed experimental data relative to the variables predicted by the calibrated model (Tedeschi, 2006). However, little attention has been paid to analytic tools that exploit the mathematical properties of the model equations. For

example, it is a typical situation to encounter difficulties when tackling model calibration due to the lack of experimental data on key system variables. Accordingly, before performing the model calibration, one might want to know if finding unique best values for the model parameters is possible given an experimental set up with specified measurements. The theoretical ability to recover the best model parameters uniquely is called structural (*a priori*) identifiability of parameters (Bellman and Astrom, 1970). Structural identifiability is a prerequisite for ensuring that the model calibration problem is well-posed (that it is a problem whose solution is unique). This property is only based on the model structure and is independent of the accuracy of experimental data. When the identifiability issue is addressed by taking into account the type and quality of available data, we refer to practical (*a posteriori*) identifiability. This paper is centred on the question of structural identifiability for models described by ODEs in animal science. Structural identifiability has been largely addressed by the community of control engineering and system identification (Walter and Pronzato, 1997). In animal science, identifiability analysis has been addressed to analyse statistical models focused on animal breeding and genetics (Wu *et al.*, 2010). With respect to dynamic ODE models, although the notion of structural identifiability was already introduced to the community (Boston *et al.*, 2007), identifiability analysis has been rarely addressed (we found only one reference of a mastitis transmission model (White *et al.*, 2002)). The lack of pervasiveness of identifiability analysis is also found in other domains of biological modelling (Roper *et al.*, 2010; Chis *et al.*, 2011b). A possible explanation of this situation is that very often identifiability analysis turns out to be difficult and demands expert knowledge on mathematical technicalities. Within this context, in this paper we address the structural identifiability analysis from a practitioner perspective by capitalizing on the use of dedicated software tools. Our

objectives are (i) to explain simply the notion of structural identifiability for the community of animal science modelling, (ii) to assess its relevance in this context and (iii) to motivate the community to the use of identifiability analysis in its modelling practice (when the concept is relevant). We want to emphasize that it is not our intention to impose on the modelling community a requirement to perform systematically identifiability analysis in their model developments. Instead, we want to open a window towards the discovery of a powerful tool for modelling construction and experiment design.

For the sake of clarity, in Table 1, we define the terms to be used in what follows. We focus mainly on dynamic models, although many aspects of what will be discussed are generic. For illustration purposes, we will use as a work-horse mathematical models describing lactation in cattle. When needed, alternative models will be used to tackle specific scenarios.

The paper is organized as follows. Firstly, a brief theoretical background on parameter identification and structural identifiability will be presented. The relevance of structural identifiability analysis will be further discussed by case studies. Aspects on practical identifiability and optimal experiment design will then be briefly addressed. Finally, the main conclusions of the work will be summarized.

Table 1 *Definition of terms used in the manuscript*

Term	Definition
System	In the modelling context, a system is a conceptual abstraction and simplification of the object under study (reality). A system consists of a set of inter-related components that interact and react as a whole to external or internal stimuli (Spedding, 1988). The system is delimited by spatial and temporal boundaries. The definition of a system sets the basis for model construction. It is of common usage to refer to the object under study as a system. Hence, we talk about system dynamics, system behaviour, etc.

Model	Set of mathematical equations derived from an abstraction and simplification of the real world (Spedding, 1988). A model is therefore a subjective formalization of knowledge on the system under study. Model construction can be motivated by two main targets: (1) understanding the functions of the system and (2) predicting the response of a set of variables for a given set of inputs. When the modelling target is that of understanding system functioning, model construction intends to describe at least partly the mechanisms that underlie the behaviour of the system under study by describing some individual elements of the system and their mutual inter-relation. In this case, the resulting model is referred to as a *mechanistic model*. A compartmental model describing the set of reactions in a metabolic pathway is an example of mechanistic model. When a model allows prediction of the time trajectories of a set of variables, the model is referred to as a *dynamic model*. Dynamic models are often described by ordinary differential equations (ODE). These models are, by construction, mostly mechanistic.
State variables	An ODE model is often referred to as a state-space model. It consists of equations describing the derivatives of characterizing variables with respect to time. These variables are called state variables. They represent the memory that the system has of its past (Khalil, 2000). They also represent systems properties such as a substrate concentration or an organ's weight (France, 1988).
Observables	Subset of the predicted variables of a model that, with respect to a defined experimental setting, can be observed (measured). In a dynamic model, they can be state variables (*e.g.,* body weight in an animal model) or a function of the state variables (*e.g.,* the pH in a rumen fermentation model). In ODE models, observables are often referred to as model outputs.
Inputs	In the dynamic modelling context, inputs are forcing variables (stimuli or challenges) external to the system that influences the system dynamics. For example, in a

	mathematical model of animal digestion, the food intake rate can be a model input.
Parameters	Scalars (assumed here to be constant) that allow the evaluation of the functions that describe the model equations. The parameters may have known values (*e.g.* physical constants such as the Avogadro number), or may need to be estimated from experimental data via model calibration.
Model structure	The model structure refers to the set of mathematical functions that specify the coupling between the state variables, the inputs and observables (Bellman and Astrom, 1970). A structural property is derived from the model equations and is (almost) independent of the values of the parameters (Walter and Pronzato, 1997). The linearity/nonlinearity of a model with respect to its parameters is an example of a structural property.
Model complexity	Throughout the manuscript, model complexity refers to the high-dimensionality of a model in terms of its parameters and state variables. Additionally, complexity is also related to the model structure: at the same number of state variables and parameters, a nonlinear model is more complex than a linear model.
Model calibration	The action of using a mathematical (numerical) routine for finding the value of unknown parameters of a model that best fit an experimental data set. The problem of finding the model parameters (an inverse problem) is formulated as the minimization of an adequate measure of the distance between the model observables and the experimental data. Model calibration is also called parameter identification (or estimation) and model fitting.
Over-parameterization	Development of models that contain more parameters than are needed to adequately describe the responses observed (Baranyi *et al.*, 1996).

Theoretical framework

Model calibration

Model calibration is the step that connects the model with the system under study.

Once experimental data on the system are available and a model structure has been

defined, the calibration (parameter identification) translates into an optimization problem, namely that of finding a set of parameters that best fits the variables predicted by the model to the data. It should be said that defining a model structure is a challenging task that represents the core of the modelling building process. Figure 1 displays a possible scheme of the parameter identification process.

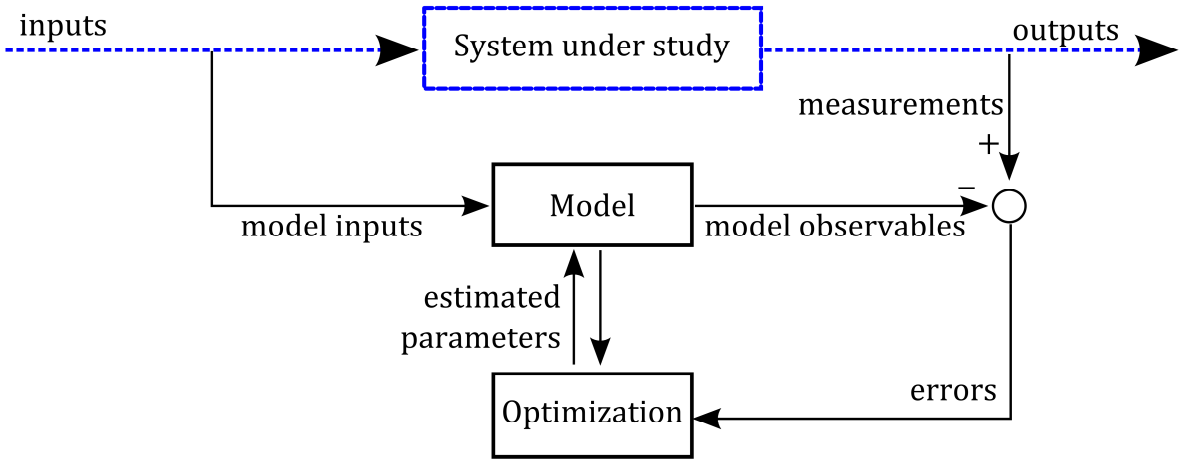

Figure 1 Scheme of the parameter identification process. A model structure has been defined to represent the dynamics of a system under study. Dashed lines represent the system (real world) and solid lines represent the virtual mathematical/numerical world. By an experimental protocol, dynamic measurements of some quantities characterizing the system behaviour have been collected. The model parameters are identified by an optimization algorithm that minimizes the model errors (distance between the measured quantities and the model observables).

Here, we consider a dynamic system subjected to an external forcing variable (input), from which we have built a mathematical model. From this system, we have collected measurements at different times of some quantities that characterize the behaviour of the system. We aim to minimize the distance between the measured quantities and their corresponding model predicted variables (observables). That is, we aim for the error (the difference between measurements and observables) to be minimum in some sense. After defining a cost function of the error (*e.g.*, the least square error), the calibration consists of adjusting the model parameters by an optimization algorithm that minimizes the defined cost function. There are a wealth of software

packages for tackling the parameter identification problem (Maiwald and Timmer, 2008; Muñoz-Tamayo *et al.*, 2009; Balsa-Canto and Banga, 2011) .

The parameter identification problem is often an ill-posed problem (it is a problem whose solution is not unique). This characteristic is the result of the different aspects, namely model structure, experimental data and numerical algorithms (Walter and Pronzato, 1997; Vargas-Villamil and Tedeschi, 2014). Ideally, we expect that the problem solution provides reliable numerical values of the parameters. In the following, we discussed tools for tackling the parameter identification problem.

Structural identifiability

Once the structure of a model is fixed and before attempting a numerical estimation of the model parameters, we might want to know if we have chances of succeeding in estimating unique optimal values of the model parameters from a given experimental setup. As previously mentioned, the possibility of recovering uniquely the model parameters relates to the mathematical property of structural identifiability, which is addressed on the sole basis of the model structure within a hypothetical ideal experiment determined by a setting of model inputs (stimuli) and observable variables (measurements). In this theoretical framework, it is assumed that the model represents perfectly the system, the observables are noise-free, and the inputs can be chosen freely to provide a sufficient excitation on the model response.

The property of structural identifiability is independent of real experimental data and is determined as follows. Let $\mathbf{M}(\mathbf{p})$ be a fixed model structure with a set of parameters $\mathbf{p} = (p_1, \cdots, p_{n_\mathrm{p}})$. $\mathbf{M}(\mathbf{p})$ describes the relationship between input variables and observables. Let us denote by $\mathbf{M}(\mathbf{p}) = \mathbf{M}(\mathbf{p}^*)$ the equality of the input-output behavior of the model structure obtained for the two parameter sets $\mathbf{p}, \mathbf{p}^*$. A

parameter p_i $(i = 1, \cdots n_{\mathrm{p}})$ is structurally identifiable if the equality $\mathbf{M}(\mathbf{p}) = \mathbf{M}(\mathbf{p}^*)$ implies that $p_i = p_i^*$, that is

$$\mathbf{M}(\mathbf{p}) = \mathbf{M}(\mathbf{p}^*) \Rightarrow p_i = p_i^* \quad (1)$$

To perform the analysis of identifiability, the equality $\mathbf{M}(\mathbf{p}) = \mathbf{M}(\mathbf{p}^*)$ is translated into a set of equations in $\mathbf{p}$**.** These equations can often be put in the form of a set of polynomial equations in $\mathbf{p}$ (parameterized by $\mathbf{p}^*$). If the resulting set of equations has a unique solution for the parameter p_i, the parameter is said to be structurally globally identifiable. If the number of solutions for p_i is finite, the parameter is structurally locally identifiable. If infinite solutions exist for p_i, the parameter is nonidentifiable. A model is structurally globally (or locally) identifiable if all its parameters are structurally globally (or locally) identifiable. A model is nonidentifiable if at least one of its parameters is nonidentifiable. A mathematical rigorous definition of structural identifiability is given by Walter and Pronzato, 1997.

Different mathematical methods exist for testing the structural identifiability of dynamic models. The tools involved include the Laplace transform, Taylor series, generating series, similarity transformation, and differential algebra. The interested reader is referred to the dedicated literature (Carson *et al.*, 1983; Walter and Pronzato, 1996; Chis *et al.*, 2011b; Raue *et al.*, 2014). In Supplementary material S1, the Laplace transform, Taylor series expansion, and generating series methods are described.

To illustrate the notion of structural identifiability, consider the following model:

$y = a \cdot b \cdot x$. We assume a hypothetical experimental protocol where x, y are measured. It is straightforward to conclude that only the quantity $a \cdot b$ is uniquely identifiable, while the individual parameters a, b are nonidentifiable. Nonidentifiability

might imply that the model is over-parameterized (see Table 1). In this trivial example, it is clear that the model can be defined by one parameter instead of two.

As mentioned in the Introduction, testing the identifiability of a model might turn out to be difficult, demanding expertise on mathematical technicalities (see Supplementary material S1). It is not our objective to go into the details of such technicalities. Rather, we take a practitioner perspective capitalizing on the developments of several software tools. These tools facilitate identifiability analysis by the practitioner (who does not have necessarily extensive knowledge in identifiability theory). Some of the identifiability software are:

DAISY (Differential Algebra for Identifiability of SYstems) (Bellu *et al.*, 2007) which is implemented in the symbolic language REDUCE, GenSSI (Generating Series for testing Structural Identifiability) (Chis *et al.*, 2011a) implemented in Matlab®, and the IdentifiabilityAnalysis application (Karlsson *et al.*, 2012) implemented in Mathematica. All of these three toolbox are freely available. The identifiability methods used by DAISY and GenSSI are explicitly referred in their acronyms. The IdentifiabilityAnalysis application uses the exact arithmetic rank approach. While DAISY and GenSSI perform global identifiability analysis, IdentifiabilityAnalysis performs local identifiability analysis, but has the advantage of allowing the analysis of complex models. Overall, the outcome of these toolboxes is a qualitative report that displays the parameters that are identifiable.

The relevance of identifiability

In the following, we discuss the relevance of the identifiability question by means of five case studies with different modelling objectives.

Case study 1: we would like to know if we have a chance of succeeding in estimating uniquely the parameters of our model

The aim pursued here is of a mathematical nature. We want to know if the parameter identification problem is well-posed. Let us consider the mathematical model proposed by Wood, 1967 that describes the lactation curve in cattle. We will refer to this model as $\mathbf{M}_W$. In this model, the daily milk production by the mammary gland (y) is described by the following gamma type algebraic equation

$$y(t) = a \cdot t^b \cdot \exp(-c \cdot t) \tag{2}$$

Where t is the time after calving and a, b, c are empirical parameters that determine the shape of the curve. $\mathbf{M}_W$ is not an ODE model (although it can be transformed into an ODE by simply deriving in time Eq. (2)). Since $\mathbf{M}_W$ is relatively simple, its identifiability can be assessed by inspection. Indeed, by taking logarithms in both sides of Eq. (2), we obtain

$$\ln y(t) = \ln a + b \cdot \ln t - c \cdot t \tag{3}$$

If continuous data of milk yield are available, it can be concluded that the model parameters are uniquely identifiable (Wood, 1967) and thus the parameter identification problem is well-posed.

Case 2: we are interested in knowing the actual value of the model parameters because of their biological relevance

In some cases, we are content with providing a model that satisfactorily predicts a variable of interest without the need to address identifiability issues. However, if our modelling objective goes beyond the purely predictive scope and we aim to improve the understanding of the phenomena that govern the system under study, the situation changes. In this case, we can be interested in knowing the actual values of

the model parameters. In mechanistic models, the parameters are biologically meaningful and we may wish to identify them uniquely because of their relevance.

Let us consider the lactation curve model proposed by Dijkstra *et al.*, 1997, that we will refer to as $\mathbf{M}_D$. In contrast to $\mathbf{M}_W$, $\mathbf{M}_D$ was originally formulated as an ODE model. Both models have equivalent predictive capabilities (Friggens *et al.*, 1999). In $\mathbf{M}_D$, the daily milk production is described by

$$\frac{\mathrm{d}y}{\mathrm{d}t} = k_1 \cdot \exp\left(-k_2 \cdot t\right) \cdot y - k_3 \cdot y, \qquad y(0) = y_0 \tag{4}$$

with y_0 the initial condition of milk production. The parameter k_1 is the specific rate of secretory cell proliferation at parturition, k_2 is a decay parameter that modulates the proliferation of secretory cells, and k_3 is a specific rate of cell death. The analytical solution of $\mathbf{M}_D$ is

$$y(t) = y_0 \cdot \exp\left\{\frac{k_1}{k_2} \cdot \left[1 - \exp\left(-k_2 \cdot t\right)\right] - k_3 \cdot t\right\} \tag{5}$$

For models with parameters that are biologically meaningful, the question of identifiability appears relevant and useful since knowing the actual value of the parameter can be of help for providing biological insight on the system under study. For example, we may wish to know unequivocally the specific rate of secretory cell proliferation at parturition (parameter k_1) of $\mathbf{M}_D$. For that, we tested the identifiability of $\mathbf{M}_D$ in Eq. (4) with the DAISY toolbox. DAISY handles models described by polynomial equations. Since $\mathbf{M}_D$ has an exponential equation, the model was suitably manipulated to facilitate the identifiability analysis as follows. We include a new state variable $x_1(t) = \exp\left(-k_2 \cdot t\right)$, which results in the following ODEs

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}t} &= k_1 \cdot x_1 \cdot y - k_3 \cdot y, \qquad y(0) = y_0 \\ \frac{\mathrm{d}x_1}{\mathrm{d}t} &= -k_2 \cdot x_1, \;\; x_1(0) = 1 \end{aligned} \tag{6}$$

If the milk yield is measured, the model parameters are uniquely identifiable. The computation time for the identifiability testing was less than one second on an Intel processor of 3.20 GHz with 8.0 GB RAM. The output file provided by DAISY is displayed in Table 2.

Table 2 *Output file of DAISY resulting from the identifiability analysis of the lactation model of Dijkstra et al., 1997. The model was suitably manipulated to be expressed with polynomial equations (a requirement of DAISY). The model is structurally globally identifiable since the basis* `gi` *provides a unique solution for all the parameters.*

```
NUMBER OF EQUATIONS$
n_ := 3$
VARIABLES VECTOR$
b_ := {y1,x1,x2}$
UNKNOWN PARAMETER(S) VECTOR$
b1_ := {k1,k2,k3}$
NUMBER OF INPUT(S)$
nu_ := 0$
NUMBER OF OUTPUT(S)$
ny_ := 1$
NUMBER OF STATE(S) $
nx_ := 2$
MODEL EQUATION(S)$
c_ := {df(x1,t)=k1*x1*x2 - k3*x1,
df(x2,t)= - k2*x2,y1=x1}$
CHARACTERISTIC SET$
aa_(1) := df(y1,t,2)*y1 - df(y1,t)**2 + df(y1,t)*y1*k2 +
y1**2*k2*k3$
aa_(2) :=  - x1 + y1$
aa_(3) := df(y1,t) - x2*y1*k1 + y1*k3
UNKNOWN PARAMETER(S) VECTOR$
b1i_ := {k1,k2,k3}$
RANDOMLY CHOSEN NUMERICAL PARAMETER(S ) VECTOR$
b2i_ := {k1=108,k2=111,k3=55}$
EXHAUSTIVE SUMMARY$
flist1i_ := {k2 - 111,k2*k3 - 6105,ic1*( - k1 + k3 + 53)}$
gi_ := {{k1=108,k3=55,k2=111}}$
MODEL GLOBALLY IDENTIFIABLE
```

To enlarge the discussion about the cases where identifiability is relevant, we tackled in Supplementary material S2, the identifiability analysis of a kinetic model of ruminal lipolysis and biohydrogentation under *in vitro* conditions (Moate *et al.*, 2008). This model has the potential to be used as primary scaffold for improving the mechanistic

description of rumen fermentation in existing models where lipid metabolism is either represented in a simplified fashion (Baldwin *et al.*, 1987; Mills *et al.*, 2001) or not accounted for (Muñoz-Tamayo *et al.*, 2016).

Case 3: the model should predict unobserved variables.

The theoretical framework of structural identifiability assumes perfect experimental data (noise-free and continuous in time). Additionally, some hypotheses on the initial conditions of the state variables need to be assumed. Initial conditions can be assumed to be unknown or fixed. These assumptions have implications on the results of identifiability testing (Saccomani *et al.*, 2003). We will use an example borrowed from Balsa-Canto and Banga, 2010, and further discussed by Villaverde and Barreiro, 2016. Let us consider a system described perfectly by the following ODE model

$$\begin{aligned} \frac{\mathrm{d}x_1(t)}{\mathrm{d}t} &= p_1 \cdot x_1 \cdot x_2, \quad x_1(0) = x_{10} \\ \frac{\mathrm{d}x_2(t)}{\mathrm{d}t} &= p_2 \cdot u, \quad x_2(0) = x_{20} \\ y_1(t) &= x_1(t) \end{aligned} \tag{7}$$

The model has two state variables (x_1, x_2). The input variable u is assumed to be known. Only the state variable x_1 can be measured. This condition is represented in the definition of the observable variable y_1. The initial conditions are set by a hypothetical experimental protocol. By solving analytically the model equations in Eq. (7), we obtain the following equation for the model observable

$$y_1(t) = x_{10} \cdot \exp(p_1 \cdot x_{20} \cdot t + 0.5 \cdot p_1 \cdot p_2 \cdot u \cdot t^2) \tag{8}$$

Let us now assume that the initial conditions are set to $x_{10} = 1, x_{20} = 0$, which leads to $y_1(t) = \exp(0.5 \cdot p_1 \cdot p_2 \cdot u \cdot t^2)$. It is clear that under these initial conditions, only

the quantity $p_1 \cdot p_2$ can be recovered from the observable variable. Hence, we can conclude that the model is nonidentifiable (*i.e.*, p_1, p_2 cannot be uniquely identified). This result implies that when performing the calibration, infinite solutions can be found which will make the calibration difficult.

Let us now assume that the model has the following *true* parameter values: $p_1 = 1.0, p_2 = 2.0$ and that the input is $u = 1.0$. By *true* parameters we refer to the ideal assumption that the model represents perfectly the system. In reality the *true* parameters are unknown. Now, under the hypothetical experimental conditions, any set of parameters fulfilling the condition $p_1 \cdot p_2 = 2.0$ is a solution of the parameter identification problem. Assume that noise-free data is available and that the optimization routine led to the following estimated parameters $p_1 = 2.0, p_2 = 1.0$. Note that the parameters fulfil the relationship $p_1 \cdot p_2 = 2.0$. To demonstrate the relevance of structural identifiability we compare the time series of x_1 and x_2 from the model simulation using the *true* parameters and the set of estimated parameters. This comparison was performed by using the original initial conditions of the hypothetical experimental protocol ($x_{10} = 1$, $x_{20} = 0$, Figure 2A) and an additional set of initial conditions ($x_{10} = 1$, $x_{20} = 0.5$, Figure 2B).

From this example, the following conclusions can be drawn:

a. If our modelling objective is to predict the dynamics of x_1, we can think at first sight that the identifiability question is irrelevant because whatever estimated parameters we obtain, we will be able to predict x_1. This reasoning, however, needs to be taken with caution. If we constrained our modelling scope to the experimental protocol with initial conditions $x_{10} = 1, x_{20} = 0$, the question of identifiability is indeed irrelevant. As observed in the left top plot of Figure 2 (panel A), the response of the two models evaluated with the *true* and estimated set of parameters are identical and thus we will

be content in finding a set of parameters such that $p_1 \cdot p_2 = 2.0$. However, if we are interested in enlarging the prediction capabilities of the model to a broader experimental context, the question of identifiability becomes relevant and necessary. We observe in the right top plot of Figure 2 (panel B) that the model response with estimated parameters differs from the response with the *true* parameters. This result implies that any prediction of x_1 in a different experimental context from that used for the model calibration will be wrong.

b. If, in addition to predicting the dynamics of x_1, we are interested in predicting x_2, the identifiability question is of greater relevance. The two bottom plots illustrate that if structural identifiability cannot be guaranteed, the model predictions of x_2 will certainly be wrong.

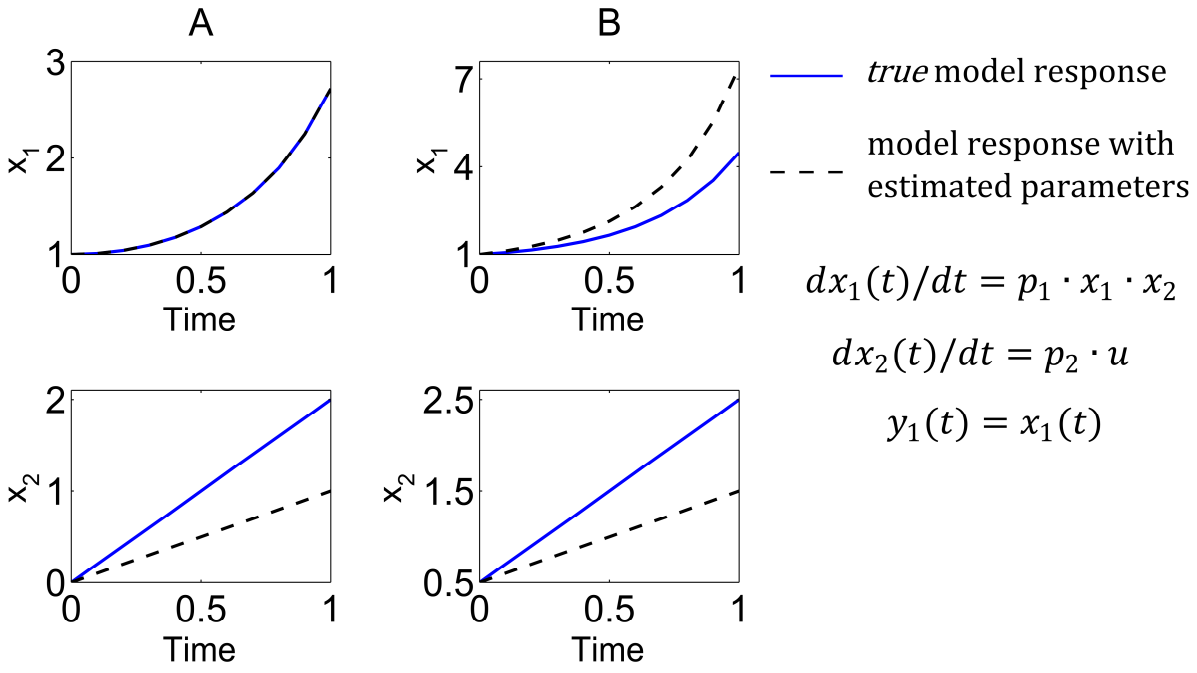

Figure 2 Relevance of structural identifiability analysis. The model response with the *true* values of parameters $p_1 = 1.0, p_2 = 2.0$ (continuous blue lines) is compared to the model response with parameters ($p_1 = 2.0, p_2 = 1.0$) obtained from a hypothetic calibration scenario (dashed black lines) with initial conditions $x_{10} = 1$, $x_{20} = 0$ (panel A), and where only x_1 can be measured ideally (noise free). Under these conditions, the model is nonidentifiable (only $p_1 \cdot p_2$ is identifiable). In panel B, the initial conditions are $x_{10} = 1, x_{20} = 0.5$. The parameters estimated from the experimental conditions in panel A cannot provide accurate predictions under other experimental conditions.

The lack of identifiability of this model can easily be reversed. Indeed, by simply setting $x_{20} > 0$, the parameters p_1, p_2 are uniquely identifiable. In Supplementary material S1, we show the structural identifiability analysis of this model by using the Taylor series and the generating series methods. We also analysed the parameter identifiability of the model using DAISY. The identifiability analysis was performed in less than one second. This example stresses the importance of the design of an experimental protocol (including the initial conditions) for guaranteeing structural identifiability.

Case 4: we attempt to use our model for testing hypotheses that cannot be verified experimentally

In animal science, the lack of experimental data on key variables can lead to multiple model structures for representing the same process (Sauvant, 1994). This multiplicity comes from the subjective nature of modelling construction that makes modelling similar to a form of art (Barnes, 1995). One of the powerful applications of mathematical modelling is that of providing a mean to address questions that are difficult to tackle experimentally. These applications include the opportunity of modelling abstract/theoretical variables that cannot be measured. We will illustrate this powerful role of models with the topic of nutrient partitioning. This issue is central in animal nutrition since the amount of nutrient that fuels a function (such as growth or lactation) is the basis for predicting nutrient requirements and develop feeding systems and recommendations.

Nutrient partitioning is regulated by two systems: a short-term system, namely homeostatic system, and a long-term system, namely homeorhetic system (Sauvant,

1994). Homeostasis regulations consists of an ensemble of adaptation and survival functions of an individual, such as glycaemia regulation after a meal. Homeorhetic regulations correspond to the orchestrated hormonal changes that drive metabolism to support the succession of physiological states that favour species survival. An example of a homeorhetic regulation is the increase of body reserves mobilisation in early-lactation to support milk production.

Different approaches exist to represent homeorhetic control of nutrient partitioning (Friggens *et al.*, 2013) but their common feature is the use of theoretical components to account for complex underlying mechanisms. These theoretical variables are used in models as proxy for translating the effects of mechanisms at underlying levels of organisation. For instance, the concept of “theoretical hormones” or “meta-hormones” has been used to represent the driving forces of body reserves changes (Hanigan *et al.*, 2007). The concept of “priorities” for life functions has been used to investigate dynamic trajectories of lactating ruminants (Puillet *et al.*, 2008; Martin and Sauvant, 2010). All these conceptual elements are used to represent the result of complex mechanisms that control nutrient partitioning and that are not possible to measure experimentally. The incorporation of theoretical driving forces in animal science modelling has been useful to move forward in predicting animal responses to their nutritional environment; *i.e.* coordinated responses of both body reserves and milk production in the dairy goat (Puillet *et al.*, 2008) and in the dairy cow (Martin and Sauvant, 2010).

Let us consider in some detail the compartmental model developed by Puillet *et al.*, 2008 to represent a homeorhetic regulatory system that controls body reserves changes and milk production across parity in dairy goats. Model equations, based on a system of priorities, are:

$$\frac{\mathrm{d}A(t)}{\mathrm{d}t} = -k_1 \cdot A, \qquad A(0) = A_0$$

$$\frac{\mathrm{d}B(t)}{\mathrm{d}t} = k_1 \cdot A - k_2 \cdot B, \qquad B(0) = B_0$$

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = k_2 \cdot B, \quad C(0) = C_0 \tag{9}$$

where the state variables A, B, C represent respectively the priorities for body reserves mobilization, milk production, and body reserves reconstitution. Simple mass-action kinetics (determined by the kinetic parameters k_1, k_2) are used to capture the major phases of body reserves changes throughout lactation process, represented as a transfer of priorities. Figure 3 displays the model schematics. At parturition, the priorities for using reserves and for producing milk are high. Then priority for body reserves mobilization decreases and, simultaneously, priority for milk production increases until it reaches a peak. This is followed by a shift in priority from milk production to body reserves reconstitution. The model structure has been constructed on biological basis. Indeed, the priority A follows an analogous dynamics to the body lipid mobilization dynamics (which can be indirectly assessed by plasma non-esterified fatty acids content), and the priority B follows an analogous dynamics to the observed dynamics of a lactation curve.

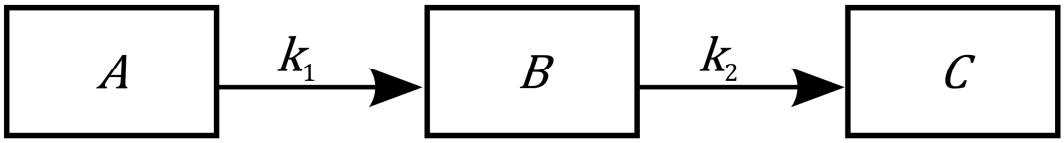

Figure 3 Schematics of the homeorhetic regulatory model of a dairy goat of Puillet *et al.*, 2008. The compartments A, B, C are respectively the priorities for body reserves mobilization, milk production, and body reserves reconstitution. System dynamics is driven by mass-action kinetics (with the parameters k_1, k_2).

We tested the identifiability of the model in Eq. (9) with DAISY. The model parameters (k_1, k_2) are uniquely identifiable if at least two state variables are measured simultaneously. They are also uniquely identifiable if either B or C are

measured and the initial conditions are known. If only A is measured, the model is nonidentifiable. The computation time for identifiability testing was less than one second.

It should be noted that the priorities described by the model are abstract variables that cannot be measured. In this case, the model construction is motivated by providing a conceptual and pertinent structure that concretizes biological hypothesis rather than producing a quantitative prediction tool. It was also constructed to overcome the existing difficulty of performing experiments to quantify homeorhetic mechanisms. As a consequence, the question of structural identifiability is in this case not relevant and does not preclude the models usefulness as a tool for understanding. Indeed, model simulations have provided useful information to analyse theoretical dynamics of phenotypic variables of interest such as milk production and body reserves.

With an academic motivation, let us analyse the hypothetical case where the priorities A, B, C of the model in Eq. (9) can be measured by an adequate experimental technique. This hypothetical case is here assumed to demonstrate the relevance that identifiability analysis can have for guiding experiment design. Suppose we plan to perform a series of experiments for estimating the model parameters with a limited budget of 10 €. The cost of measuring $A, B,$ and C is respectively 2 €, 8 €, and 10 €. How to select what to measure? Well, given the critical situation of funding in research, we will be tempted to choose to measure only A and use the remaining 8 € in other projects. This decision is of course wrong, because measuring only A will not provide quality information for model calibration. Measuring either B or C will be adequate. If we measure only B, we will have 2 € to

compensate our financial deficit. However, if we want to get the most out of the experiment, the wisest choice is to measure both A and B.

Let us now consider the following model and assume that it can be an alternative model of the regulatory model in Eq. (9)

$$\frac{\mathrm{d}A(t)}{\mathrm{d}t} = -\frac{k_1}{k_3 + A} \cdot A \cdot B, \qquad A(0) = A_0$$

$$\frac{\mathrm{d}B(t)}{\mathrm{d}t} = \frac{k_1}{k_3 + A} \cdot A \cdot B - k_2 \cdot B, \qquad B(0) = B_0$$

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = k_2 \cdot B, \quad C(0) = C_0 \tag{10}$$

This model is more complex than the model in Eq. (9). Firstly, it is nonlinear because the flux from A to B is described by a nonlinear function (Michaelis-Menten kinetics) instead of a first-order kinetics, and secondly it has one additional parameter (k_3). We tested the identifiability of the model using DAISY, GenSSI and IdentifiabilityAnalysis. The computation time was less than one second in all the three toolboxes. In this case, the parameters of the model are identifiable if any of the state variables is measured. If only A is measured the model parameters are identifiable, which contrasts to the identifiability properties of the original model described by Eq. (9) (nonidentifiable if only A is measured). The result appears at first sight as counterintuitive, since in modelling practice complex models are often penalized. In the framework of structural identifiability, nonlinear models tend to be more identifiable than linear models (Walter and Pronzato 1996; Roper *et al.*, 2010). In the previous example, adding a nonlinearity and a supplementary parameter help to improve the structural identifiability of the model. However, the increase of the number of parameters has, in general, a negative influence on the practical identifiability by rendering the model calibration harder, and increasing the risk of overfitting.

Case 5: what if in our modelling scenario the question of structural identifiability is relevant but our model is nonidentifiable?

As it was previously mentioned, the lack of experimental data on key variables imposes a particular challenge in the modelling task and can lead to various difficulties including the lack of structural identifiability. Models where the number of parameters is very high with respect to number of observables may lack of structural identifiability. Although, as demonstrated in the case study 4, we cannot affirm systemically that models with more parameters are less identifiable than models with less parameters. The identifiability depends on the model structure and on how the parameters appear in the observables. After this clarification, the lack of identifiability of mathematical models is not an uncommon scenario, and is often encountered in domains such as system biology. Then, what to do? One popular solution consists in capitalizing on existing knowledge by setting some parameters to known values reported in the literature. This strategy results in reducing the number of unknown parameters to be estimated and may favour the identifiability of the reduced parameter set. Caution should be paid in selecting parameters obtained from experimental conditions that are compatible to the case study. Parameter reduction can also be performed by grouping some of the model parameters (Schaber and Klipp, 2011). A second solution is to design a new experiment that renders the model identifiable by selecting an adequate set of observables (Anguelova *et al.*, 2012).

If after exhausting the above-mentioned alternatives the nonidentifaibility cannot be eliminated, this does not necessarily mean that our model is useless. Firstly, the modelling construction requires the verbal hypotheses on the system under study to become specific and conceptually rigorous (Schaber and Klipp, 2011). This

conceptual step is central for gaining insight on the system behaviour, and pointing out the aspects that need to be deepened. Secondly, if our modelling purpose is to predict, we can assess numerically to what extend the lack of identifiability can impact model predictions. We can identify which set of variables of the model are the less sensitive to the actual values of the parameters (Gutenkunst *et al.*, 2007) and which model predictions can be uniquely determined despite lack of identifiability (Cedersund, 2012). We emphasised that the nonidentifiability of a model does not preclude its usefulness. A relevant example is the model of the circadian clock in *Arabidopsis thaliana* (Locke *et al.*, 2005). The model has 29 parameters, from which only 17 parameters are at least locally identifiable under certain stimuli (Chis *et al.*, 2011b). Although its lack of identifiability, the model of *Arabidopsis thaliana* represents an important modelling contribution for enhancing understanding of the loops of genes that drive circadian locks in living organisms. By being aware of the lack of identifiability and by using adequate tools, nonidentifiable models can still be useful by providing both qualitative and quantitative information for gaining insight on system behaviour (Schaber and Klipp, 2011; Cedersund, 2012).

A brief summary about the relevance of identifiability discussed here is given in Table 3.

Table 3 *Summary of the case studies for assessment the relevance of structural identifiability*

Case study	Model objective	Scientific question addressed	Relevance of identifiability
1	To represent an observed variable for further prediction	Is the parameter estimation well-posed?	Yes
2	To represent an observed variable by using a biologically based model for further prediction	Can we theoretically know the actual value of a parameter that is biologically meaningful?	Yes
3	To represent an observed variable and predict an unobserved variable	Can we guarantee highly quality predictions for variables that cannot be experimentally	Yes

		measured?	
4	To provide a conceptual modelling framework of phenomena that are difficult to evaluate experimentally	Is it a model consisting of abstract variables that cannot be experimentally measured biologically pertinent?	No
5	To represent mechanistically a complex biological process	Is it a nonidentifiable model useful?	Yes

To sum up, structural identifiability analysis remains desirable whenever feasible in the model calibration context, since it determines if the parameter identification problem has a unique solution when we have unlimited available data. Identifiability testing can provide guidelines for designing experiments and be useful for facilitating model simplification by identifying some potential over-parameterizations. Together, this information facilitates the model calibration step. Identifiability analysis is therefore relevant when (i) we are interested in the actual values of the model parameters, and (ii) we want to predict variables, in particular those that cannot be measured directly. Although structural identifiability is a desired property of a model, we clearly state that a nonidentifiable model can still be a useful model.

In addition to the identifiability methods previously mentioned, identifiability analysis can also be performed numerically, for example by using interval analysis (Braems *et al.*, 2001). Numerical approaches allow the computational complexity associated with identifiability methods based on algebraic manipulations of observable derivatives to be overcome. Indeed, when dealing with complex models, identifiability analysis may be impossible to perform even with the advanced software tools applied here, and recent developments (Villaverde *et al.*, 2016), and thus the assessment of structural identifiability by numerical means is of great value. A very intuitive solution consists in using *prior* values of the model parameters for generating simulated data for a hypothetical experimental setup and perform the model calibration. By inspection, we

can assess if the resulting parameter estimates are close to the *priors* values used for data generation. If this is this case, the model might be at least locally identifiable (Walter and Pronzato, 1997). A more sophisticated solution is that of the profile likelihood approach (Raue *et al.*, 2009) which provides a powerful numerical method for assessing structural and practical identifiability of high-dimension models.

On practical identifiability and optimal experiment design for parameter estimation

In the previous section, we mentioned that structural identifiability is a necessary condition for the well-posedness of the model calibration problem. However, structural identifiability does not guarantee the accuracy of the estimation and the quality of the model predictions (Carson *et al.*, 1983). In practice, we aim to find accurate parameter estimates from experimental data. The actual accuracy of the parameter identification depends on the characteristics of the actual experimental data. The question to be addressed is: for a fixed model structure and given a set of experimental data, how accurate will be the estimated parameters? Data are always corrupted by noise, and are usually in short supply (although this situation is rapidly changing due to the progress of precision farming technologies). Hence, even if the model is structurally identifiable, the quality of the estimation can be poor, leading to parameter estimates that can even take values that are physically meaningless. Furthermore, there might exist many sets of parameter values that fit the data equally well, which can be troublesome for drawing biological based conclusions as discussed by Boer *et al.*, 2017 when addressing the parameter estimation of a bovine estrous cycle model.

Tackling the accuracy of the parameter identification with respect to experimental data is the core of practical identifiability. To illustrate the notion of practical identifiability, let us consider the model $y = a \cdot x_1 + b \cdot x_2$ and assume that the variables x_1, x_2, y can be measured. The parameters a, b are structurally identifiable. Now, consider that experimental data are available and that $x_1, x_2,$ are proportional ($i.e.\ x_2 = c \cdot x_1,$). With these data, the parameters a, b are not practically identifiable. The only quantity that is practically identifiable is $a + b \cdot c$. The parameter estimates under these experimental conditions will not be accurate. Accuracy of the parameter estimation is related to parameter uncertainty (high accuracy implies low uncertainty) and is assessed by the computation of the confidence intervals of the parameter estimates. Large confidence intervals imply low reliability on the parameter estimates (practical unidentifiability).

One classical approach for determining the confidence intervals of the parameter estimates is via the computation of the Fisher Information Matrix (FIM). In Supplementary material S3, we recall the principles of this classical approach and introduce some aspects of optimal experiment design (OED) for parameter estimation. The goal of OED is to find, under a set of constraints, an experiment setup that allows an accurate estimation of the model parameters (which translates in small confidence intervals). To illustrate the power of OED, consider the curve lactation model $\mathbf{M}_D$ in Eq. (4). The OED problem is defined with a *prior* nominal parameter set (extracted from literature or experimental data). Let us assume that these nominal values are $k_1 = 0.1, k_2 = 0.15, k_3 = 0.005$, and that the initial condition of yield milk is y_0 =10 with t_0 = 0.1 d. We aim to find three sampling times (along 100 days) that provide high informative content for estimating the model parameters accurately. For that, we defined an OED problem in which the optimal sampling times

were found by maximizing the determinant of the FIM. Maximizing the determinant of the FIM implies minimizing the volume of the confidence intervals (see Supplementary material S3). The FIM was obtained by symbolic manipulation using the Matlab® Toolbox IDEAS (Muñoz-Tamayo *et al.*, 2009), which is freely available at http://genome.jouy.inra.fr/logiciels/IDEAS. The OED problem was solved using the Nelder–Mead Simplex method implemented in Matlab®.

The optimal sampling times were: t_0 = 0.1 d, t_1 = 7.5 d, t_2 = 29 d, t_3 = 100 d. For comparison, we calculated the confidence intervals obtained from an equidistant sampling time setup (t_0 = 0.1 d, t_1 = 33.4 d, t_2 = 66.7 d, t_3 = 100 d). Figure 4 displays the obtained optimal sampling times, together with the sampling times from the equidistant strategy.

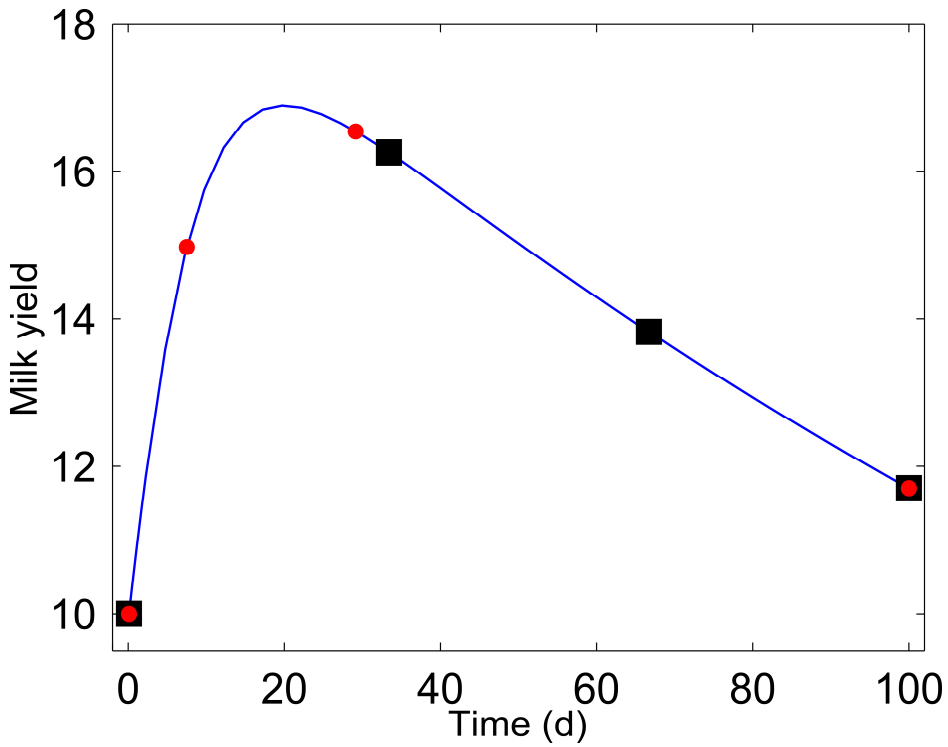

Figure 4 Lactation model of Dijkstra *et al.*, 1997. Equidistant sampling times (■), vs sampling times obtained from optimal experiment design (●).

Table 4 shows the comparative results. Optimal sampling times obtained from OED provide substantially a better accuracy of the estimation than equidistant sampling

times. For k_1, k_2, the standard deviations from the OED are only 5% of the standard deviations provided by the equidistant sampling setup. For k_3, the standard deviation from the OED is 50% of that obtained with the equidistant sampling setup.

Table 4 *Accuracy of the parameter estimates of the lactation model of Dijkstra et al., 1997 for an optimal sampling strategy and a equidistant sampling strategy*

	Accuracy of the estimation $(\pm 2 \cdot s.d.)$	
Parameter	Equidistant strategy	OED strategy
k_1	0.035	0.0014
k_2	0.055	0.0028
k_3	0.0002	0.0001

This example illustrates the capabilities of OED and the interest of incorporating this tool into our modelling practice when data have not been collected yet. OED allows maximum exploitation of experimental data for model calibration, and avoid pitfalls from applying traditional experiment designs without cautious analysis. In fact, it is common practice to use factorial designs for defining an experimental setup. If the levels are not chosen adequately, the factorial design can lead to practical identifiability problems such a singular FIM (see Muñoz-Tamayo *et al.*, 2014 for an illustrative example). If this occurs, the reliability of the parameter estimates cannot be assessed given that confidence intervals computation requires the FIM to be invertible (see Supplementary material 3).

It goes without saying that the identification of model parameters is a very challenging problem, where difficulties are encountered even for models of moderate complexity. In the case of a complex model, one may wonder however about the practical relevance of providing a result about the identifiability of the model, given that identifying the parameters of the model from actual noisy data is already extremely difficult. In this respect, by studying the parameter sensitivities of a collection of 17 models of biological systems, Gutenkunst *et al.*, 2007 have elaborated the concept of sloppiness, that establishes that some parameters (sloppy)

can change by orders of magnitude without affecting significantly the model output. Sloppiness is related to the condition number of the FIM (Supplementary material S3) and results from high differences between the eigenvalues of the FIM. The parameter identification of a sloppy model data suffers from high uncertainty as a result of a singular (ill-conditioned) FIM. The authors claimed that sloppiness is a universal property of systems biology models and suggest that modellers should focus on predictions rather than on identifying the actual values of the model parameters. Given this work, it may seem tempting to desist from any efforts to look for an accurate parameter identification. Nevertheless, the notion of sloppiness has been subject of debate and its value as conceptual tool has been questioned. In the comprehensive work of Chis *et al.*, 2016, it has been demonstrated that sloppy models can be identifiable and that OED can substantially improve the practical identifiability of models, even if they are complex. Chis *et al.*, 2016 suggested that OED should be performed on the basis of classical criteria such as maximizing the determinant of the FIM instead of looking at minimizing model sloppiness. Accordingly, addressing parameter identifiability in complex models is not a hopeless quest when the adequate tools are deployed.

Conclusions

This article was centred on introducing and discussing the mathematical tool of structural identifiability analysis, which has been seldom applied in animal science modelling. This lack of pervasiveness in our domain is probably due to the mathematical technicalities which identifiability analysis relies on. These technicalities are beyond the academic background in animal science. But this hurdle can be overcome by adopting a practitioner perspective and capitalizing on existing

dedicated identifiability software that should facilitate the application of identifiability analysis in our domain. By using illustrative examples, we attempted to open a window towards the discovery of a powerful tool for modelling construction and experiment design when the identifiability question is relevant. Overall, identifiability analysis is relevant when the purpose of the modelling construction is the prediction of variables that cannot be measured, and when we are interested in knowing the actual value of the model parameters. Finally, the success to getting the most out of structural identifiability analysis in animal science modelling relies on a constructive dialog between experimenters and modellers.

Acknowledgements

We would like to thank the developers of the identifiability toolboxes DAISY (Bellu *et al.*, 2007), GenSSI (Chis *et al.*, 2011a) and IdentifiabilityAnalysis (Karlsson *et al.*, 2012) for rendering their software freely available and favouring Open Science. Many thanks to Dr. Maria Pia Saccomani (University of Padova, Italy) for helpful insights on the use of DAISY. We thank Nicolas Friggens (INRA-AgroParisTech, France) for his critical comments on the manuscript. Rafael Muñoz-Tamayo is indebted to Éric Walter for his rigorous and provocative explanations on the fascinating subject of parameter identification.

References

Anguelova M, Karlsson J and Jirstrand M 2012. Minimal output sets for identifiability. Mathematical Biosciences 239, 139-153.

Baldwin RL 2000. Introduction: history and future of modelling nutrient utilization in farm animals. In Modelling nutrient utilization in farm animals (ed. JP McNamara, France, J., Beever, D.E.), pp. 1-9, CAB International, Wallingford, UK.

Baldwin RL, Thornley JH and Beever DE 1987. Metabolism of the lactating cow. II. Digestive elements of a mechanistic model. J. Dairy Res. 54, 107-131.

Balsa-Canto E and Banga JR 2010. Advanced model identification using global optimization. Theoretical introduction. http://www.iim.csic.es/~amigo/ICSB2010_tutorial_AMIGO_theory.pdf.

Balsa-Canto E and Banga JR 2011. AMIGO, a toolbox for advanced model identification in systems biology using global optimization. Bioinformatics 27, 2311-2313.
Baranyi J, Ross T, McMeekin TA and Roberts TA 1996. Effects of parameterization on the performance of empirical models used in 'predictive microbiology'. Food Microbiology 13, 83-91.
Barnes CJ 1995. The Art of Catchment Modeling - What Is a Good Model. Environment International 21, 747-751.
Bellman R and Astrom KJ 1970. On structural identifiability. Math. Biosci. 7, 329-339.
Bellu G, Saccomani MP, Audoly S and D'Angio L 2007. DAISY: A new software tool to test global identifiability of biological and physiological systems. Computer Methods and Programs in Biomedicine 88, 52-61.
Boer HM, Butler ST, Stotzel C, Te Pas MF, Veerkamp RF and Woelders H 2017. Validation of a mathematical model of the bovine estrous cycle for cows with different estrous cycle characteristics. Animal, 1-11.
Boston RC, Wilkins P and Tedeschi LO 2007. Identifiability and Accuracy: Two critical problems associated with the application of models in nutrition and the health sciences. In Mathematical modeling for nutrition and health sciences (ed. M Hanigan), pp. 161-193, Roanoke, VA, USA.
Braems L, Jaulin L, Kieffer M and Walter E 2001. Guaranteed numerical alternatives to structural identiflability testing. Proceedings of the 40th Ieee Conference on Decision and Control, Vols 1-5, 3122-3127.
Carson ER, Cobelli C and Finkelstein L 1983. The Mathematical Modeling of Metabolic and Endocrine Systems: Model Formulation, Identification, and Validation. John Wiley & Sons, New York.
Cedersund G 2012. Conclusions via unique predictions obtained despite unidentifiability - new definitions and a general method. Febs Journal 279, 3513-3527.
Chis O-T, Villaverde AF, Banga JR and Balsa-Canto E 2016. On the relationship between sloppiness and identifiability. Mathematical Biosciences 282, 141-161.
Chis O, Banga JR and Balsa-Canto E 2011a. GenSSI: a software toolbox for structural identifiability analysis of biological models. Bioinformatics 27, 2610-2611.
Chis OT, Banga JR and Balsa-Canto E 2011b. Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods. PLoS One 6.
Dijkstra J, France J, Dhanoa MS, Maas JA, Hanigan MD, Rook AJ and Beever DE 1997. A model to describe growth patterns of the mammary gland during pregnancy and lactation. Journal of Dairy Science 80, 2340-2354.
Doeschl-Wilson AB 2011. The role of mathematical models of host-pathogen interactions for livestock health and production - a review. Animal 5, 895-910.
France J 1988. Mathematical-Modeling in Agricultural Science. Weed Research 28, 419-423.
Friggens NC, Emmans GC and Veerkamp RF 1999. On the use of simple ratios between lactation curve coefficients to describe parity effects on milk production. Livestock Production Science 62, 1-13.
Friggens NC, Brun-Lafleur L, Faverdin P, Sauvant D and Martin O 2013. Advances in predicting nutrient partitioning in the dairy cow: recognizing the central role of genotype and its expression through time. Animal 7, 89-101.
Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR and Sethna JP 2007. Universally sloppy parameter sensitivities in systems biology models. PLoS Comput Biol 3, 1871-1878.
Hanigan MD, Rius AG, Kolver ES and Palliser CC 2007. A redefinition of the representation of mammary cells and enzyme activities in a lactating dairy cow model. Journal of Dairy Science 90, 3816-3830.
Karlsson J, Anguelova M and Jirstrand M 2012. An efficient method for structural identifiability analysis of large dynamic systems. In 16th IFAC Symposium on System Identification pp. 941-946.
Khalil HK 2000. Nonlinear systems. Pearson Education International Inc, Upper Saddle River, New Jersey

Locke JCW, Millar AJ and Turner MS 2005. Modelling genetic networks with noisy and varied experimental data: the circadian clock in Arabidopsis thaliana. Journal of Theoretical Biology 234, 383-393.
Maiwald T and Timmer J 2008. Dynamical modeling and multi-experiment fitting with PottersWheel. Bioinformatics 24, 2037-2043.
Martin O and Sauvant D 2010. A teleonomic model describing performance (body, milk and intake) during growth and over repeated reproductive cycles throughout the lifespan of dairy cattle. 1. Trajectories of life function priorities and genetic scaling. Animal 4, 2030-2047.
Mills JA, Dijkstra J, Bannink A, Cammell SB, Kebreab E and France J 2001. A mechanistic model of whole-tract digestion and methanogenesis in the lactating dairy cow: model development, evaluation, and application. J. Anim. Sci. 79, 1584-1597.
Moate PJ, Boston RC, Jenkins TC and Lean IJ 2008. Kinetics of ruminal lipolysis of triacylglycerol and biohydrogenation of long-chain fatty acids: New insights from old data. Journal of Dairy Science 91, 731-742.
Muñoz-Tamayo R, Giger-Reverdin S and Sauvant D 2016. Mechanistic modelling of in vitro fermentation by rumen microbiota. Anim. Feed Sci. Technol. 220, 1-21.
Muñoz-Tamayo R, Laroche B, Leclerc M and Walter E 2009. IDEAS: a Parameter Identification Toolbox with Symbolic Analysis of Uncertainty and its Application to Biological Modelling. In Prepints of the 15th IFAC Symposium on System Identification, Saint-Malo, France, pp. 1271-1276.
Muñoz-Tamayo R, Martinon P, Bougaran G, Mairet F and Bernard O 2014. Getting the most out of it: Optimal experiments for parameter estimation of microalgae growth models. Journal of Process Control 24, 991-1001.
Puillet L, Martin O, Tichit M and Sauvant D 2008. Simple representation of physiological regulations in a model of lactating female: application to the dairy goat. Animal 2, 235-246.
Raue A, Karlsson J, Saccomani MP, Jirstrand M and Timmer J 2014. Comparison of approaches for parameter identifiability analysis of biological systems. Bioinformatics 30, 1440-1448.
Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmuller U and Timmer J 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25, 1923-1929.
Roper RT, Saccomani MP and Vicini P 2010. Cellular signaling identifiability analysis: A case study. Journal of Theoretical Biology 264, 528-537.
Saccomani MP, Audoly S and D'Angio L 2003. Parameter identifiability of nonlinear systems: the role of initial conditions. Automatica 39, 619-632.
Sauvant D 1994. Modeling Homeostatic and Homeorhetic Regulations in Lactating Animals. Livestock Production Science 39, 105-113.
Schaber J and Klipp E 2011. Model-based inference of biochemical parameters and dynamic properties of microbial signal transduction networks. Current Opinion in Biotechnology 22, 109-116.
Spedding CRW 1988. General aspects of modelling and its application in livestock production. In Modelling of livestock production systems (eds. S Korver and JAM Van Arendonk), pp. 3-13, Kluwer Academica Publishers, Dordrecht, The Netherlands
Tedeschi LO 2006. Assessment of the adequacy of mathematical models. Agricultural Systems 89, 225-247.
Vargas-Villamil LM and Tedeschi LO 2014. Potential integration of multi-fitting, inverse problem and mechanistic modelling approaches to applied research in animal science: a review. Animal Production Science 54, 1905-1913.
Villaverde AF and Barreiro A 2016. Identifiability of large nonlinear biochemical networks. MATCH Commun. Math. Comput. Chem 76, 259-296.
Villaverde AF, Barreiro A and Papachristodoulou A 2016. Structural Identifiability of Dynamic Systems Biology Models. Plos Computational Biology 12.

Walter E and Pronzato L 1996. On the identifiability and distinguishability of nonlinear parametric models. Mathematics and Computers in Simulation 42, 125-134.
Walter E and Pronzato L 1997. Identification of Parametric Models from Experimental Data. Springer, London.
White LJ, Evans ND, Lam TJGM, Schukken YH, Medley GF, Godfrey KR and Chappell MJ 2002. The structural identifiability and parameter estimation of a multispecies model for the transmission of mastitis in dairy cows with postmilking teat disinfection. Mathematical Biosciences 180, 275-291.
Wood PDP 1967. Algebraic Model of Lactation Curve in Cattle. Nature 216, 164-&.
Wu XL, Heringstad B and Gianola D 2010. Bayesian structural equation models for inferring relationships between phenotypes: a review of methodology, identifiability, and applications. Journal of Animal Breeding and Genetics 127, 3-15.

Supplementary material S1

Three methods for performing structural identifiability analysis of dynamic models

This section describes briefly three methods for testing structural identifiability in dynamic models. Consider the model described by the following ordinary differential equations

$$\begin{aligned} \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} &= \mathbf{f}(\mathbf{x},\mathbf{u},\mathbf{p},t), \qquad \mathbf{x}(\mathbf{0}) = \mathbf{x_0} \\ \mathbf{y_m}(t) &= \mathbf{g}(\mathbf{x},\mathbf{u},\mathbf{p},t) \end{aligned} \tag{1}$$

where t is the time, $\mathbf{x}$ is the vector of state variables, $\mathbf{y_m}$ is the vector of model observables, and $\mathbf{u}$ is the vector of external stimuli (input vector). The equations contain a set of parameters defined by the vector $\mathbf{p}$, and $\mathbf{f}$, $\mathbf{g}$ are vector functions.

Laplace Transform

If the model in Eq. (1) is linear, a classical approach for testing its structural identifiability is via the analysis of the transfer function of the model resulting from the Laplace transformation (Bellman and Astrom, 1970). The transfer function matrix $\mathbf{H}(s,\mathbf{p})$ is defined by

$$\mathbf{H}(s,\mathbf{p}) = \frac{\mathbf{Y}(s,\mathbf{p})}{\mathbf{U}(s,\mathbf{p})} \tag{2}$$

where s is the argument of the Laplace domain, $\mathbf{Y}(s,\mathbf{p})$ and $\mathbf{U}(s,\mathbf{p})$ are the Laplace transforms of the observables ($\mathbf{y_m}$) and inputs ($\mathbf{u}$).

Once $\mathbf{H}(s,\mathbf{p})$ is written in canonical form, we can proceed to write the transfer function matrix for two parameters sets $\mathbf{p},\mathbf{p}^*$. Further, by establishing the relation $\mathbf{H}(s,\mathbf{p}) \equiv \mathbf{H}(s,\mathbf{p}^*)$ we can derive a set of equations translating the identities of the coefficients of $\mathbf{H}(s,\mathbf{p})$ and $\mathbf{H}(s,\mathbf{p}^*)$.

If the solution for the set of equations is unique for $\mathbf{p}$, that is $\mathbf{p} = \mathbf{p}^*$, the model is structurally identifiable.

For illustration, let us consider the following single-input and single-output (SISO) model

$$\begin{aligned} \frac{\mathrm{d}x(t)}{\mathrm{d}t} &= (a+b)\cdot x + c\cdot u, \quad x_0 = 0 \\ y(t) &= x(t) \end{aligned} \tag{3}$$

with parameters a,b,c and the input u. The observable $y(t)$ is the state variable $x(t)$. By applying the Laplace transform, we obtain

$$sX(s) = (a + b) \cdot X(s) + c \cdot U(s) \tag{4}$$

where $X(s), U(s)$ correspond respectively to the state variable and the input variable in the Laplace domain. The model observable in the Laplace domain is $Y(s) = X(s)$. The transfer function is given by

$$H(s) = \frac{Y(s)}{U(s)} = \frac{c}{s - (a + b)} \tag{5}$$

The identity equations are

$$c = c^* \tag{6}$$

$$a + b = a^* + b^* \tag{7}$$

From Eq. (6) and Eq. (7), we can conclude that the parameter c is uniquely identifiable while the parameters a, b are nonidentifiable since Eq. (7) have infinite solutions.

Many examples of identifiability analysis for linear compartmental models are presented in Carson *et al.*, 1983.

Taylor series expansion

This approach was developed by Pohjanpalo, 1978. It assumes that the vector functions **f, g** in Eq. (1) are continuously differentiable in their arguments, implying that the state and the observable vectors can have infinitely many time derivatives. The development of the Taylor series of the observable $\mathbf{y_m}(t)$ in the model described by Eq. (1) results

$$\mathbf{y_m}(t) = \mathbf{y_m}(0) + t\frac{\mathrm{d}\mathbf{y_m}}{\mathrm{d}t}(0) + \frac{t^2}{2!}\frac{\mathrm{d}^2\mathbf{y_m}}{\mathrm{d}t^2}(0) + \cdots + \frac{t^k}{k!}\frac{\mathrm{d}^k\mathbf{y_m}}{\mathrm{d}t^k}(0), k = 0,1,2,\cdots,\infty \tag{8}$$

Let us denote

$$a_k = \frac{\mathrm{d}^k\mathbf{y_m}}{\mathrm{d}t^k}(0) \tag{9}$$

Since the observable vector is a unique function of time, all its derivatives (a_k) are unique and known. The structural identifiability of the model is determined from the analysis of the equations of the successive derivatives a_k evaluated at two parameters sets $\mathbf{pp}^*$. The model is structurally identifiable if

$$\mathbf{a}_k(\mathbf{p}) = \mathbf{a}_k(\mathbf{p}^*),\ k = 0,1,2,\cdots,k_{\max} \Rightarrow\ \mathbf{p} = \mathbf{p}^* \tag{10}$$

where $k_{\max}$ is at least the number of unknown parameters.

As example, consider the following model

$$\frac{\mathrm{d}x_1(t)}{\mathrm{d}t} = p_1 \cdot x_1 \cdot x_2, \quad x_{10} = 1.0$$

$$\frac{\mathrm{d}x_2(t)}{\mathrm{d}t} = p_2 \cdot u, \quad x_{20} = 2.0$$

$$y_1(t) = x_1(t) \tag{11}$$

With parameters p_1, p_2 and the input u. The model has two state variables $x_1(t), x_2(t)$ and one observable $y_1(t)$ that corresponds to the state variable $x_1(t)$. By developing the successive derivatives of $y_1(t)$, we obtain

$$a_0 = x_{10} = 1.0 \tag{12}$$

$$a_1 = p_1 \cdot x_{10} \cdot x_{20} = 2.0 \cdot p_1 \tag{13}$$

$$a_2 = p_1^2 \cdot x_{10} \cdot x_{20}^2 + p_1 \cdot p_2 \cdot x_{10} \cdot u = 4.0 \cdot p_1^2 + p_1 \cdot p_2 \cdot u \tag{14}$$

The model is globally identifiable. The parameter p_1 can be uniquely obtained from the coefficient a_1, and subsequently p_2 can be uniquely recovered from a_2.

Generating series

This method was developed by Walter and Lecourtier, 1982 and it is conceptually similar to the Taylor series approach. Consider the model described by the following ordinary differential equations

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{f}_0(\mathbf{x}, \mathbf{u}, \mathbf{p}, t) + \sum_{i=1}^{m} \mathbf{f}_i(\mathbf{x}, \mathbf{p}, t)\mathbf{u}_i \ , \qquad \mathbf{x}(\mathbf{0}) = \mathbf{x_0}$$
$$\mathbf{y_m}(t) = \mathbf{g}(\mathbf{x}, \mathbf{p}, t) \tag{15}$$

where $\mathbf{f}_i$ $(i = 0,1,\cdots,m)$ and $\mathbf{g}$ are analytic, implying that the model observables can be expanded in series with respect to time and the model inputs. The coefficients of the series are $\mathbf{g}(\mathbf{x}(t), \mathbf{p}, t)$ and the successive Lie derivatives evaluated at $t = 0$

$$\boldsymbol{L}_{\mathbf{f}_{j_0}} \cdots \boldsymbol{L}_{\mathbf{f}_{j_k}} \mathbf{g}(\mathbf{x}, \mathbf{p}, t)\Big|_0 \tag{16}$$

where $\boldsymbol{L}_{\mathbf{f}}\mathbf{g}(\mathbf{x}, \mathbf{p}, t)$ is the Lie derivative of $\mathbf{g}$ along $\mathbf{f}$, defined by

$$\boldsymbol{L}_{\mathbf{f}}\mathbf{g}(\mathbf{x}, \mathbf{p}, t) = \sum_{i=1}^{n_x} \mathbf{f}_i(\mathbf{x}, \mathbf{p}, t) \frac{\partial \mathbf{g}(\mathbf{x}, \mathbf{p}, t)}{\partial \mathbf{x}_j} \tag{17}$$

with $\mathrm{n_x}$ the number of state variables.

Analogous to the Taylor series, let $\mathbf{s}(\mathbf{p})$ the vector of the series coefficients. The model is structurally identifiable if (Walter and Pronzato, 1996).

$$\mathbf{s}(\hat{\mathbf{p}}) = \mathbf{s}(\mathbf{p}^*) \ \Rightarrow \hat{\mathbf{p}} = \mathbf{p}^* \tag{18}$$

As example, consider again the model in Eq. (11), which can be written as

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \begin{bmatrix} p_1 \cdot x_1 \cdot x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ p_2 \end{bmatrix} u \ , \qquad \mathbf{x}(\mathbf{0}) = \begin{bmatrix} 1.0 & 2.0 \end{bmatrix}^{\mathrm{T}}$$
$$\mathbf{y_m}(t) = x_1 \tag{19}$$

where $\mathbf{f}_0(\mathbf{x}, \mathbf{u}, \mathbf{p}, t) = \begin{bmatrix} p_1 \cdot x_1 \cdot x_2 \\ 0 \end{bmatrix}$, $\mathbf{f}_1(\mathbf{x}, \mathbf{u}, \mathbf{p}, t) = \begin{bmatrix} 0 \\ p_2 \end{bmatrix}$, $\mathbf{g}(\mathbf{x}, \mathbf{p}, t) = x_1$

The first Lie derivative operators are

$$\boldsymbol{L}_{\mathbf{f}_0} = [p_1 \cdot x_1 \cdot x_2] \frac{\partial}{\partial x_1} \tag{20}$$

$$\boldsymbol{L}_{\mathbf{f}_1} = p_2 \frac{\partial}{\partial x_2} \tag{21}$$

The coefficients of the series are the following Lie derivatives

$$\boldsymbol{L}_{\mathbf{f}_0}\mathbf{g}|_0 = [p_1 \cdot x_1 \cdot x_2] \frac{\partial x_1}{\partial x_1}\Big|_0$$

$$= \ p_1 \cdot x_1 \cdot x_2|_0 = 2 \cdot p_1 \tag{22}$$

$$\boldsymbol{L}_{\mathbf{f}_1}\boldsymbol{L}_{\mathbf{f}_0}\mathbf{g}|_0 = p_2 \frac{\partial}{\partial x_2}\{\boldsymbol{L}_{\mathbf{f}_0}\mathbf{g}\}\Big|_0$$

$$= \ p_2 \frac{\partial}{\partial x_2}\{p_1 \cdot x_1 \cdot x_2\}|_0$$

$$= p_1 \cdot p_2 \cdot x_1|_0 = p_1 \cdot p_2 \tag{23}$$

From the coefficient in Eq. (22), it is deduced that p_1 is identifiable. From Eq. (23), we obtain that p_2 is identifiable.

Finally, the interested reader is referred to recent literature on structural identifiability methods and their comparison (Chis *et al.*, 2011; Miao *et al.*, 2011; Raue *et al.*, 2014).

References

Bellman R and Astrom KJ 1970. On structural identifiability. Math. Biosci. 7, 329-339.

Carson ER, Cobelli C and Finkelstein L 1983. The Mathematical Modeling of Metabolic and Endocrine Systems: Model Formulation, Identification, and Validation. John Wiley & Sons, New York.

Chis OT, Banga JR and Balsa-Canto E 2011. Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods. PLoS One 6.

Miao HY, Xia XH, Perelson AS and Wu HL 2011. On Identifiability of Nonlinear ODE Models and Applications in Viral Dynamics. Siam Review 53, 3-39.

Pohjanpalo H 1978. System identifiability based on the power series expansion of the solution. Mathematical Biosciences 41, 21-33.

Raue A, Karlsson J, Saccomani MP, Jirstrand M and Timmer J 2014. Comparison of approaches for parameter identifiability analysis of biological systems. Bioinformatics 30, 1440-1448.

Walter E and Lecourtier Y 1982. Global Approaches to Identifiability Testing for Linear and Non-Linear State-Space Models. Mathematics and Computers in Simulation 24, 472-482.
Walter E and Pronzato L 1996. On the identifiability and distinguishability of nonlinear parametric models. Mathematics and Computers in Simulation 42, 125-134.

Supplementary material S2

Structural identifiability of a ruminal lipolysis and biohydrogenation model

To enlarge our discussion on structural identifiability, we analyse here two mathematical models developed by Moate *et al.*, 2008 to represent *in vitro* kinetics of two biological processes namely ruminal lipolysis and biohydrogentation. These two multistep biological processes were described mathematically by using a multi-compartmental modelling approach. Each model was calibrated by published experimental data (Noble *et al.*, 1974) where only the biological process of interest took place. Reactions rates were defined by either Michaelis-Menten kinetics or first-order kinetics.
For the lipolysis model, the set of ordinary differential equations of the model are

$$\frac{\mathrm{d}x_1(t)}{\mathrm{d}t} = -\frac{k_1 \cdot x_1}{k_2 + x_1} \cdot e^{-k_3 \cdot t}, \qquad x_1(0) = x_{10}$$

$$\frac{\mathrm{d}x_2(t)}{\mathrm{d}t} = \frac{2}{3} \cdot \frac{k_1 \cdot x_1}{k_2 + x_1} \cdot e^{-k_3 \cdot t} - k_4 \cdot x_2, \;\; x_2(0) = 0$$

$$\frac{\mathrm{d}x_3(t)}{\mathrm{d}t} = \frac{1}{2} \cdot k_4 \cdot x_2 - k_4 \cdot x_3, \qquad x_3(0) = 0$$

$$\frac{\mathrm{d}x_4(t)}{\mathrm{d}t} = \frac{1}{3} \cdot \frac{k_1 \cdot x_1}{k_2 + x_1} \cdot e^{-k_3 \cdot t} + \frac{1}{2} \cdot k_4 \cdot x_2 + k_4 \cdot x_3, \;\; x_4(0) = 0 \qquad (1)$$

where x_1 is the concentration of triglyceride fatty acids, x_2 is the concentration of diglyceride fatty acids , x_3 is the concentration of monoglyceride fatty acids and x_4 is the concentration of non-esterified fatty acids. From the experimental setup, the model observables are

$$y_1(t) = x_1(t)$$

$$y_2(t) = x_2(t) + x_3(t)$$

$$y_3(t) = x_4(t) \qquad (2)$$

The observable $y_2(t)$ is an aggregated pool of mono and diglyceride fatty acids named by Noble *et al.*, 1974 as partial glycerides.

Identifiability analysis was performed using DAISY (Bellu *et al.*, 2007), GenSSI (Chis *et al.*, 2011) and IdentifiabilityAnalysis (Karlsson *et al.*, 2012). The model parameters are identifiable. For all of these software tools, the computation time was less than one second on an Intel processor of 3.20 GHz with 8.0 GB RAM.
Interestingly, we determined that the observable $y_2(t) = x_2(t) + x_3(t)$ was not necessary to guarantee the identifiability of the model parameters. This result indicates that sometimes having many observations does not imply necessarily an improvement on the structural identifiability of a model. Indeed, in a context of resource-consuming measurements, we might be interested in identifying a minimal set of measurements that guarantee structural identifiability (Anguelova *et al.*, 2012).

However, it is clear that in practice having more measurements can be instrumental for performing model calibration.

For the biohydrogenation model, the set of ordinary differential equations of the model are

$$\frac{\mathrm{d}x_4(t)}{\mathrm{d}t} = -\frac{k_5 \cdot x_4}{k_6 + x_4}, \qquad x_4(0) = x_{40}$$

$$\frac{\mathrm{d}x_5(t)}{\mathrm{d}t} = \frac{k_5 \cdot x_4}{k_6 + x_4} - \frac{k_7 \cdot x_5}{k_8 + x_5 + x_6}, \quad x_5(0) = 0$$

$$\frac{\mathrm{d}x_6(t)}{\mathrm{d}t} = \frac{k_7 \cdot x_5}{k_8 + x_5 + x_6} - k_9 \cdot x_6 \frac{k_{10} - x_6}{k_{10}}, \qquad x_6(0) = \ 0$$

$$\frac{\mathrm{d}x_7(t)}{\mathrm{d}t} = k_9 \cdot x_6 \frac{k_{10} - x_6}{k_{10}}, \ x_7(0) = \ 0 \tag{3}$$

where x_4 is the concentration of non-esterified fatty acids (linoleic), x_5 is the concentration of rumenic acid, x_6 is the concentration of vaccenic and x_7 is the concentration of stearic acid. From the experimental setup, the model observables are

$$y_1(t) = \ x_4(t)$$

$$y_2(t) = \ x_5(t)$$

$$y_3(t) = \ x_6(t)$$

$$y_4(t) = \ x_7(t) \tag{4}$$

Structural identifiability analysis was performed with DAISY. The model parameters are identifiable. The computation time was less than one second.

It should be noted however, that the accuracy of the parameter estimates strongly depends on the quality of the available data for calibration. Indeed, Moate *et al.*, 2008 encountered practical identifiability problems for estimating some model parameters. To circumvent this obstacle, some parameters were fixed and set as known values.

Note that a mathematical model representing both lipolysis and biohydrogenation should integrate the model equations in Eq. (1) and Eq. (3). For the non-esterified fatty acids (x_4), the resulting differential equation is

$$\frac{\mathrm{d}x_4(t)}{\mathrm{d}t} = \frac{1}{3} \cdot \frac{k_1 \cdot x_1}{k_2 + x_1} \cdot e^{-k_3 \cdot t} + \frac{1}{2} \cdot k_4 \cdot x_2 + k_4 \cdot x_3 - \frac{k_5 \cdot x_4}{k_6 + x_4}, \ x_4(0) = \ x_{40} \tag{5}$$

The full model integrating both lipolysis and biohydrogenation has seven state variables and ten parameters. As an academic exercise, we analysed the structural identifiability of the full model using the model observables in Eq. (2) and Eq. (4). The model parameters were identifiable. The computation time in DAISY was about 30

seconds, indicating the computational effort required when model complexity increases (30 seconds vs 1 second).

The results from structural identifiability analysis are encouraging for modelling attempts towards an enhanced mechanistic representation of the rumen ecosystem. Although improvements are needed (in particular for describing biohydrogenation), the model developed by (Moate *et al.*, 2008) provided a parsimonious and biological based approach that can be used as scaffold for incorporating lipid metabolism in existing models of rumen fermentation (Baldwin *et al.*, 1987; Mills *et al.*, 2001; Muñoz-Tamayo *et al.*, 2016). In a scenario of constructing a predictive model of rumen fermentation, the question of identifiability of the model of lipolysis and biohydrogenation is relevant since parameter estimates obtained from *in vitro* data can be used as priors in an extended model describing the *in vivo* system.

References

Anguelova M, Karlsson J and Jirstrand M 2012. Minimal output sets for identifiability. Mathematical Biosciences 239, 139-153.

Baldwin RL, Thornley JH and Beever DE 1987. Metabolism of the lactating cow. II. Digestive elements of a mechanistic model. J. Dairy Res. 54, 107-131.

Bellu G, Saccomani MP, Audoly S and D'Angio L 2007. DAISY: A new software tool to test global identifiability of biological and physiological systems. Computer Methods and Programs in Biomedicine 88, 52-61.

Chis O, Banga JR and Balsa-Canto E 2011. GenSSI: a software toolbox for structural identifiability analysis of biological models. Bioinformatics 27, 2610-2611.

Karlsson J, Anguelova M and Jirstrand M 2012. An efficient method for structural identifiability analysis of large dynamic systems. In 16th IFAC Symposium on System Identification pp. 941-946.

Mills JA, Dijkstra J, Bannink A, Cammell SB, Kebreab E and France J 2001. A mechanistic model of whole-tract digestion and methanogenesis in the lactating dairy cow: model development, evaluation, and application. J. Anim. Sci. 79, 1584-1597.

Moate PJ, Boston RC, Jenkins TC and Lean IJ 2008. Kinetics of ruminal lipolysis of triacylglycerol and biohydrogenation of long-chain fatty acids: New insights from old data. Journal of Dairy Science 91, 731-742.

Muñoz-Tamayo R, Giger-Reverdin S and Sauvant D 2016. Mechanistic modelling of in vitro fermentation by rumen microbiota. Anim. Feed Sci. Technol. 220, 1-21.

Noble RC, Moore JH and Harfoot CG 1974. Observations on the pattern on biohydrogenation of esterified and unesterified linoleic acid in the rumen. Br J Nutr 31, 99-108.

Supplementary material S3

Calculation of confidence interval from the Fisher Information Matrix and a brief comment on optimal experiment design

Assessment of the uncertainty of the parameter estimates

In this section, we recall the theoretical framework for assessing the uncertainty of the parameter identification by using the Fisher Information Matrix following the classic book of Walter and Pronzato, 1997.

Let us consider the following model described by ordinary differential equations

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t), \qquad \mathbf{x}(\mathbf{0}) = \mathbf{x_0}$$
$$\mathbf{y_m}(t) = \mathbf{g}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t) \tag{1}$$

where t is the time, $\mathbf{x}$ is the vector of state variables, $\mathbf{y_m}$ is the vector of model observables, and $\mathbf{u}$ is the vector of external stimuli (input vector). The equations contain a set of parameters defined by the vector $\mathbf{p}$, and $\mathbf{f}$, $\mathbf{g}$ are vector functions.

When real experimental data are available, represented here by the vector $\mathbf{y}(t)$, we can proceed to the model calibration step by finding the vector $\mathbf{p}$ that minimizes a cost function of the distance between the real measurements $\mathbf{y}(t)$ and the model observables $\mathbf{y_m}(t)$.

It is typical to assume that the vector of experimental data collected at time t_i follows

$$\mathbf{y}(t_i) = \mathbf{y_m}(t_i, \mathbf{p}^*) + \boldsymbol{\varepsilon_i}, \; i = 1,2, \dots, n_\mathrm{t} \tag{2}$$

where n_t is the number of observation times, $\mathbf{y_m}(t_i, \mathbf{p}^*)$ is the predicted observable of the model with $\mathbf{p}^*$ the true value of the parameter vector and $\boldsymbol{\varepsilon}_i$ is the vector of measurement errors which will be assumed here to follow a normal distribution: $\boldsymbol{\varepsilon}_i \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

The model calibration can be performed by the maximum likelihood (ML) approach. If the covariance matrix $\boldsymbol{\Sigma}$ is known, the maximum likelihood estimator minimizes the weighted least squares function. Once the parameters estimates ($\hat{\mathbf{p}}$) are found by an adequate optimization procedure, we can assess the parameter uncertainty via the computation of the FIM at the estimated value $\hat{\mathbf{p}}$ as detailed below.

The FIM can be calculated as

$$\mathrm{FIM}(\hat{\mathbf{p}}) = \sum_{i=1}^{n_\mathrm{t}} \left[\frac{\partial \mathbf{y_m}}{\partial \mathbf{p}}\right]^\mathrm{T}_{(t_i, \hat{\mathbf{p}})} \boldsymbol{\Sigma} \left[\frac{\partial \mathbf{y_m}}{\partial \mathbf{p}}\right]_{(t_i, \hat{\mathbf{p}})} \tag{3}$$

The term $\frac{\partial \mathbf{y_m}}{\partial \mathbf{p}}$ contains the sensitivities of the observables with respect to the parameters. The calculation of the sensitivities can be performed by symbolic manipulation of the model equations using dedicated software such as the Matlab®

Toolbox IDEAS (Muñoz-Tamayo *et al.*, 2009), which is freely available at http://genome.jouy.inra.fr/logiciels/IDEAS.

Under a number of technical assumptions that include theoretical identifiability, the covariance matrix $\mathbf{P}$ of the ML estimator satisfies

$$\mathbf{P} \geq \mathrm{FIM}^{-1}(\mathbf{p}^*) \tag{4}$$

This equation is known as the Cramér-Rao inequality. An approximate $\widehat{\mathbf{P}}$ of the covariance matrix of the parameters can be computed at the Cramér-Rao lower bond evaluated at $\widehat{\mathbf{p}}$

$$\widehat{\mathbf{P}} = \mathrm{FIM}^{-1}(\widehat{\mathbf{p}}) \tag{5}$$

It must be kept in mind that this approximation is only valid asymptotically, when the number of data points tends to infinity, the statistical hypotheses on the noise are satisfied, and that $\widehat{\mathbf{p}}$ is close to $\mathbf{p}^*$**.** Furthermore, this approach is based on a linear approximation of the observables with respect to the parameters, which may be inadequate because of model nonlinearities (Carson *et al.*, 1983, Marsili-Libelli *et al.*, 2003, Raue *et al.*, 2009). When these idealized conditions are far from being satisfied, this evaluation of the uncertainty on the estimates via the FIM has thus to be considered with caution.

The diagonals of $\widehat{\mathbf{P}}$ are the variances of the parameter estimates. Thus, the square root σ_j of the *j*th diagonal element of $\widehat{\mathbf{P}}$ is an estimate of the standard deviation of the parameter $\hat{p}_j$. On this basis, an approximate 95% confidence interval of the parameter $\hat{p}_j$ can be calculated as $\hat{p}_j \pm 2 \cdot \sigma_j$. It should be noted that the determination of the covariance matrix $\widehat{\mathbf{P}}$ requires the FIM to be invertible (nonsingular). The condition number of the FIM (*i.e.*, the ratio of the largest eigenvalue of the FIM to the smallest) is a useful indicator of the practical identifiability of the model given the available data. The higher the condition number, the more difficult the optimization is and the lower practical identifiability.

A brief introduction to optimal experiment design for parameter estimation.

When designing an experimental configuration with the aim of providing data to be used for model calibration, we expect the resulting data to be highly informative for allowing accurate estimation of the model parameters. The problem of defining such an experimental configuration is the realm of optimal experiment design (OED) for parameter estimation.

Since, the FIM is the core for determining the confidence intervals of the parameter estimates, classical approaches of OED for parameter estimation rely on the optimization of a scalar function of the FIM. The most popular criteria for OED is the D-optimality criterion, which maximizes the determinant of the FIM, and the E-optimality criterion, which maximizes the smallest eigenvalue of the FIM. Maximizing the determinant of the FIM implies minimizing the volume of the confidence ellipsoids for the parameters (Walter and Pronzato, 1997), while maximizing the smallest eigenvalue of the FIM implies minimizing the maximum diameter of the confidence ellipsoids for the parameters.

The OED problem can be formulated mathematically as follows

$$\min_{\boldsymbol{\varphi}} j\left(\mathrm{FIM}(\mathbf{p}, \boldsymbol{\varphi})\right) \tag{6}$$

where $j\left(\mathrm{FIM}(\mathbf{p}, \boldsymbol{\varphi})\right)$ is a scalar cost function of the FIM (*e.g.*, $\det\left(\mathrm{FIM}(\mathbf{p}, \boldsymbol{\varphi})\right)$) and $\boldsymbol{\varphi}$ is the design vector that defines the experimental configuration (*e.g.* sampling times, initial conditions, stimuli). Since the true values of the model parameters are unknown, the OED problem is defined with a nominal parameter set $\mathbf{p}^0$**,** whose values are obtained from literature or experimental data. This value can be further refined in an iterative process. It should be noted that the OED problem is constrained by experimental limitations and is only possible when there are some degrees of freedom in the procedure for data collection. Finally, the solution of the OED problem requires efficient optimization algorithms as discussed in the dedicated literature (Walter and Pronzato, 1997; Balsa-Canto *et al.*, 2008).

References

Balsa-Canto E, Alonso AA and Banga JR 2008. Computational procedures for optimal experimental design in biological systems. IET Syst Biol 2, 163-172.

Carson ER, Cobelli C and Finkelstein L 1983. The Mathematical Modeling of Metabolic and Endocrine Systems: Model Formulation, Identification, and Validation. John Wiley & Sons, New York.

Marsili-Libelli S, Guerrizio S and Checchi N 2003. Confidence regions of estimated parameters for ecological systems. Ecological Modelling 165, 127-146.

Muñoz-Tamayo R, Laroche B, Leclerc M and Walter E 2009. IDEAS: a Parameter Identification Toolbox with Symbolic Analysis of Uncertainty and its Application to Biological Modelling. In Prepints of the 15th IFAC Symposium on System Identification, Saint-Malo, France, pp. 1271-1276.

Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmuller U and Timmer J 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25, 1923-1929.

Walter E and Pronzato L 1997. Identification of Parametric Models from Experimental Data. Springer, London.